MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

AD-A257 259

(2)

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE Aug 92	3. REPORT TYPE AND DATES COVERED Final 1 Sep 91-31 May 92

4. TITLE AND SUBTITLE International Conference on Random Mappings, Partitions and Permutations	5. FUNDING NUMBERS DAAL03-91-G-0322
6. AUTHOR(S) Simon Tavare	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Univ. of Southern California Los Angeles, CA 90089-1113	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARO 29280.1-EL-CF

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The conference was held as scheduled. A report has been prepared which contains abstracts of papers presented at the conference.

14. SUBJECT TERMS Conference, Random Mappings, Partitions, Permutations, Random Graphs, Probability	15. NUMBER OF PAGES 46 16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

ARO 29280.1-EL-CF

International Conference on Random Mappings, Partitions and Permutations

University of Southern California

January 3 – 6, 1992

PROGRAM

ORGANIZERS:

Richard Arratia (USC)

Solomon Golomb (USC)

Bernard Harris (UW Madison)

Gian-Carlo Rota (MIT)

Simon Tavaré (USC; Chair)

GRANT SUPPORT FROM:

Air Force Office of Scientific Research

Army Research Office

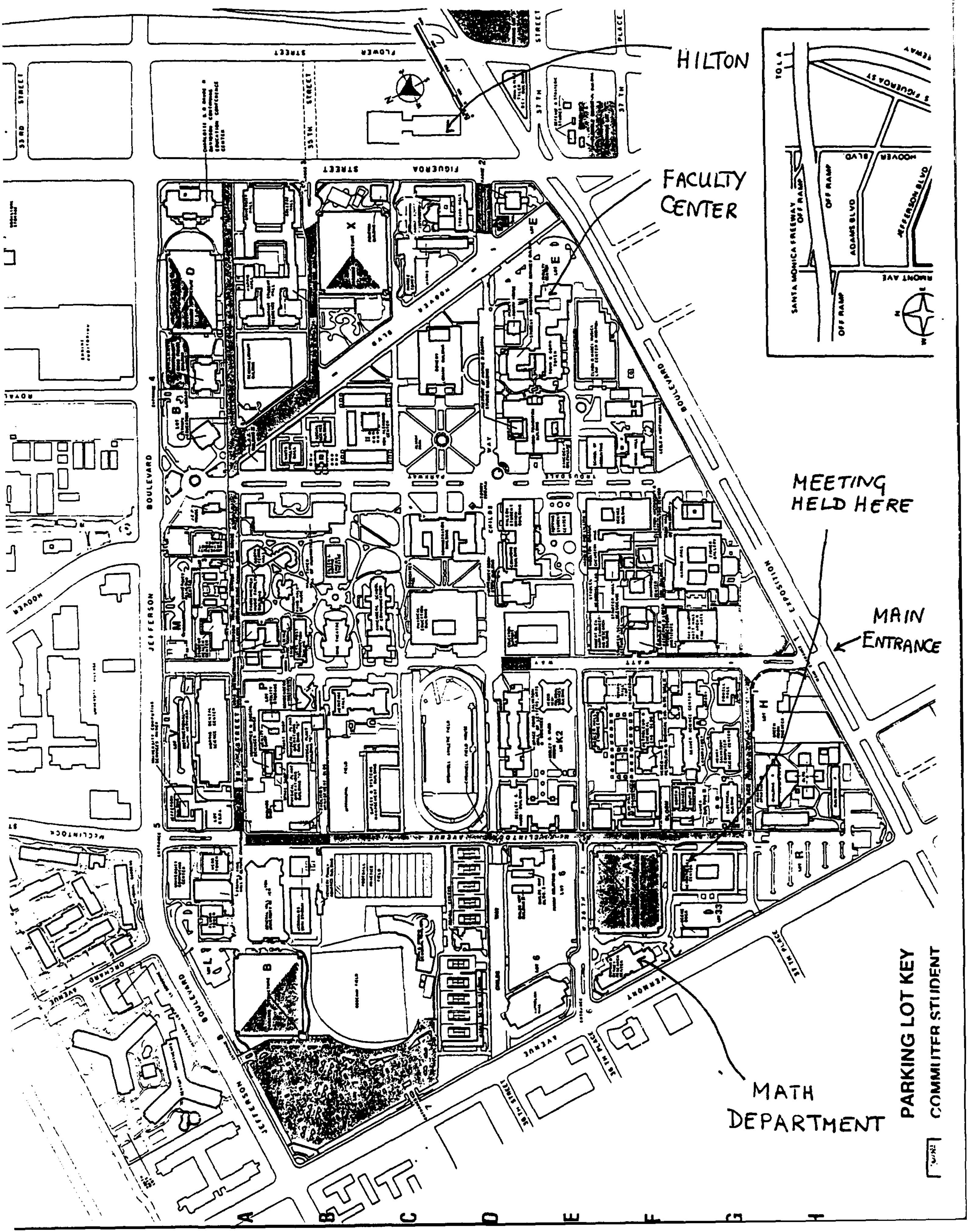
University of Southern California

National Science Foundation

National Security Agency

Office of Naval Research

The most convenient car access is at Gate 1, on Hoover and Exposition. The lectures will take place in the Davis Auditorium of the Ethel Percy Andrus Gerontology Center in the southwest corner of the USC campus. Refreshments will be served in Room 224.


HILTON
FACULTY CENTER
MEETING HELD HERE
MAIN ENTRANCE
MATH DEPARTMENT
FIGUEROA STREET
FLOWER STREET
EXPOSITION BOULEVARD
JEFFERSON BOULEVARD
VERMONT AVENUE
HOOVER BLVD
SANTA MONICA FREEWAY
ADAMS BLVD
JEFFERSON BLVD
OFF RAMP
PARKING LOT KEY
COMMUTER STUDENT

Friday, January 3

8:15 – 8:45	Refreshments
8:45 – 8:50	R.E. KAPLAN, Vice–Provost, USC *Welcome*
8:50 – 9:00	G.-C. ROTA, MIT *Introductory Remarks*
9:00 – 9:40	H.S. WILF, University of Pennsylvania *Ascending subsequences of permutations and Young tableaux*
9:40 – 10:20	B. HARRIS, University of Wisconsin, Madison *The early history of the theory of random mappings*
10:20 – 10:40	Break
10:40 – 11:40	V.F. KOLCHIN, Steklov Mathematical Institute, Moscow *Cycles in random graphs and hypergraphs*
11:40 – 1:30	Lunch
1:30 – 2:10	D.J. ALDOUS, University of California, Berkeley *Brownian bridge asymptotics for random mappings*
2:10 – 2:50	A.M. ODLYZKO, AT&T Bell Laboratories *Search for the maximum of a random walk*
2:50 – 3:10	Break
3:10 – 3:50	B. BOLLOBÁS, University of Cambridge, England *The height of a random partial order: concentration of measure*
3:50 – 4:30	P. DIACONIS, Harvard University *Comparison techniques for card shuffling*
4:30 – 4:50	Break
4:50 – 5:30	P. ERDŐS *Recent problems in probabilistic number theory and combinatorics*

Saturday, January 4

8:15 – 9:00	Refreshments
9:00 – 9:20	J.M. STEELE, University of Pennsylvania *Long common subsequence problems*
9:20 – 9:40	L. HOLST, Royal Institute of Technology, Stockholm *On menage problems*
9:40 – 10:00	P.J. JOYCE, University of Idaho *Poisson limit laws for dependent random permutations*
10:00 – 10:20	V. KHOKHLOV, Steklov Mathematical Institute, Moscow *On the structure of a non-uniformly distributed random graph*
10:20 – 10:40	Break
10:40 – 11:20	J. SPENCER, New York University *The Poisson paradigm and random graphs*
11:20 – 12:00	L. TAKÁCS, Case Western Reserve University *On the heights and widths of random rooted trees*
12:00 – 1:30	Lunch and Photo
1:30 – 2:10	A. VERSHIK, St. Petersburg University *Random permutations, limit shapes and asymptotic problems of partition theory*
2:10 – 2:50	L.H. HARPER, University of California, Riverside *In search of maximum antichains of partitions*
2:50 – 3:10	Break
3:10 – 3:30	W. STADJE, University of Osnabrück, Germany *On sets of integers with prescribed gaps*
3:30 – 3:50	E. SCHMUTZ, Drexel University *On random partitions of the integer n*
3:50 – 4:10	B. FRISTEDT, University of Minnesota *Random partitions of large integers*
6:00 – 7:00	Cash Bar (Faculty Center)
7:00 – 10:00	Conference Dinner

Sunday, January 5

8:15 – 9:00	Refreshments
9:00 – 9:20	W.J. EWENS, University of Pennsylvania *Sampling properties of random mappings*
9:20 – 9:40	D.R. GAVELEK, XonTech Inc. *The height of elements in random mappings*
9:40 – 10:00	J. JAWORSKI, Adam Mickiewicz University, Poland *The evolution of a random mapping*
10:00 – 10:20	J.C. HANSEN, Northeastern University *Order statistics for random combinatorial structures*
10:20 – 10:40	Break
10:40 – 11:20	O.V. VISKOV, Steklov Mathematical Institute, Moscow *The Rota umbral calculus and the Heisenberg–Weyl algebra*
11:20 – 12:00	B. PITTEL, Ohio State University *Random permutations and stable matchings*
12:00 – 1:30	Lunch
1:30 – 2:10	R.A. ARRATIA, University of Southern California *Independent process approximations for random combinatorial structures*
2:10 – 2:50	L.A. SHEPP, AT&T Bell Laboratories *Linear and non-linear codes for a special channel*
2:50 – 3:10	Break
3:10 – 3:30	P. MATTHEWS, University of Maryland, Baltimore County *A lower bound on the probability of conflict under non-uniform access in database systems*
3:30 – 3:50	P. TETALI, DIMACS Center, Rutgers University *Covering with Latin transversals*
3:50 – 4:10	Z.-X. HU, University of Illinois *On $([n], P)$ partitions*
4:10 – 4:30	A.P. GODBOLE, Michigan Technological University *Some results on Poisson and compound Poisson approximation*

Monday, January 6

8:15 – 9:00	Refreshments
9:00 – 9:40	P.J. DONNELLY, Queen Mary and Westfield College, London *Labellings, size-biased permutations and the GEM distribution*
9:40 – 10:20	V.A. VATUTIN, Steklov Mathematical Institute, Moscow *Branching processes with final types of particles and random trees*
10:20 – 10:40	Break
10:40 – 11:20	A.D. BARBOUR, University of Zürich, Switzerland *Refined approximations for the Ewens sampling formula*
11:20 – 12:00	J.W. PITMAN, University of California, Berkeley *Cycles and descents of random permutations*
12:00 – 12:15	Farewell
12:15 – 1:30	Lunch

Ascending subsequences of permutations and Young tableaux.

H.S. WILF

University of Pennsylvania

It is well known, from Shensted's algorithm, that there is a relationship between the longest increasing subsequence in a permutation and the length of the first row of a Young tableau. We give here a quantitative version. That is, we give an explicit relationship between the numbers of permutations of n letters whose longest increasing subsequence is of length k and of Young tableaux of n cells whose first row has length k. The relationship is surprisingly simple. A number of unsolved problems are raised.

The early history of the theory of random mappings

B. HARRIS

University of Wisconsin, Madison

The paper will review much of the early work in the theory of random mappings. The contributions of Ulam, Katz, Riordan and the speaker will be reviewed. The paper will also discuss work by Rubin and Sitgreaves, Folkert, and Lenard. This work is of particular interest, since it has never been published.

Cycles in random graphs and hypergraphs

V.F. KOLCHIN

Steklov Mathematical Institute

For a $T \times n$ matrix $A = \|a_{ij}\|$ in $GF(2)$ we define a hypergraph G_A with n vertices and T hyperedges

$$e_t = \{j : a_{ij} = 1\}, \qquad t = 1, \ldots, T.$$

Denote $a_t = (a_{t1}, \ldots, a_{tn}\}$, $t = 1, \ldots\, T$.

A set of row numbers $\{t_1, \ldots, t_n\}$ is called a critical set if the sum of vectors

$a_{t_1} + \ldots + a_{t_m}$ is the zero vector.

We can naturally define the concept of independence for critical sets and determine the maximal number $s(A)$ of independent critical sets in A. The total number of critical sets $S(A)$ is equal to $2^{s(A)} - 1$. It is not difficult to see that $s(A)$ and the rank $r(A)$ of the matrix A are connected by the equality

$$r(A) + s(A) = T.$$

Therefore we can use the more tractable characteristic $s(A)$ for the investigations of rank of the matrix A.

We consider a matrix A of a special form which corresponds to the following system of T random equations in $GF(2)$:

$$x_{i_1(t)} + \ldots + x_{i_r(t)} = b_t, \qquad t = 1, \ldots, T,$$

where $i_1(t), \ldots, i_r(t)$, $t = 1, \ldots, T$, are independent identically distributed random variables which take values $1, \ldots, n$ with equal probabilities. We denote by $A_{r,n,T}$ the matrix of this system.

In the case $r = 2$ a critical set of the matrix $A_{2,n,T}$ corresponds to an ordinary cycle in the ordinary graph $G_{A_{2,n,T}}$. The behaviour of the number of cycles in such graphs is well known [1–4]. If $n, T \to \infty$ in such a way that $2T/n \to \lambda$, $0 < \lambda < 1$, then the distribution of the number of cycles converges to the Poisson distribution with parameter $\Lambda = -\frac{1}{2}\ln(1 - \lambda)$. In [3] a new proof of this assertion is given. The number of cycles in a non-equiprobable graph is investigated in [5].

In the case $r > 2$ we introduce a concept of a hypercycle as a set of hyperedges which corresponds to a critical set of $A_{r,n,T}$. We prove a threshold property for the mean number of hypercycles in the hypergraph $G_{A_{r,n,T}}$.

Let $r \geq 3$ be fixed, $T, n \to \infty$ in such a way that $T/n \to \alpha$. Then there exists a constant α_r such that $\mathsf{M}S(A_{r,n,T}) \to 0$ if $\alpha < \alpha_r$, and $\mathsf{M}S(A_{r,n,T}) \to \infty$ if $\alpha > \alpha_r$.

The constant α_r is the first component of the vector which is the only solution of the following system of equations in three unknowns a, x, λ:

$$\begin{aligned} e^{-x}\cosh\lambda\left(\frac{ar}{ar-x}\right)^{a} &= 1, \\ \frac{x}{\lambda}\left(\frac{ar-x}{x}\right)^{1/r} &= 1, \\ \lambda\tanh\lambda &= x. \end{aligned}$$

A numerical analysis of this system gives us the following values of the critical constants:

$$\alpha_3 = 0{,}8894\ldots, \quad \alpha_4 = 0{,}9671\ldots, \quad \alpha_5 = 0{,}9891\ldots,$$

$$\alpha_6 = 0{,}9969\ldots, \quad \alpha_7 = 0{,}9986\ldots, \quad \alpha_8 = 0{,}9995\ldots.$$

REFERENCES

1. P. Erdős and A. Rényi, On the evolution of random graphs. *Publ. Math. Inst. Hungaria Acad. Sci.* (1960) **5**, 17–61.

2. L. Takács, On the limit distribution of the number of cycles in a random graph. *J. Appl. Probab.* (1988) **25A**, 359–376.

3. V. F. Kolchin, On the number of cycles in a random graph. *Probab. Problems of Discrete Math.* MIEE, Moscow, 1990, pp. 3–8.

4. V. F. Kolchin, On the behaviour of a random graph near a critical point. *Theory Probab. Appl.* (1986) **31**, 439–451.

5. V. F. Kolchin and V. I. Khokhlov, On the number of cycles in a non-equiprobable random graph. *Diskretnaya Matematika* (1990) **2**, No. 3, 137–145 (in Russian).

6. G. V. Balakin, V. F. Kolchin and V. I. Khokhlov, Hypercycles in a random hypergraph. *Diskretnaya Matematika* (1991) **3**, No. 3, 102–108 (in Russian).

Brownian bridge asymptotics for random mappings

D. Aldous

University of California, Berkeley

Write $\mathbf{n} = \{1, 2, \ldots, n\}$. A function $f : \mathbf{n} \to \mathbf{n}$ may be regarded as a directed graph with edges $i \to f(i)$; note this allows an edge $i \to i$. By a *random mapping* F_n we mean a uniform random choice of f from all n^n functions $\mathbf{n} \to \mathbf{n}$. There is a large literature on combinatorial analysis of random mappings, much due to the Soviet school. Their results up to the early 1980s can be found summarized in Kolchin [3]. Three distinct methods have classically been used for proving $n \to \infty$ asymptotics for random mappings.

- Take limits in exact formulas.
- Generating function methods: see e.g. Flajolet and Odlyzko [2].
- Representing certain quantities as i.i.d. random variables conditioned on their sum: see Kolchin [3].

and more recently Stein's method has been used to bound the errors in certain asymptotic approximations.

The purpose of this talk is to present a new method. We show how a mapping can be coded as a walk (with steps ± 1) of length $2n$. Our main result is that the random walk coded from the random mapping can be rescaled to converge as $n \to \infty$ to reflecting Brownian bridge (rBB). This one result encompasses many asymptotic results for particular statistics which have previously been treated separately – loosely, it gives limit distributions for all "global" functionals of random mappings. Of course, the limit distribution is given in terms of a corresponding functional of rBB, which requires some calculation to evaluate explicitly. Fortunately most distributions of interest have already been discussed in the theoretical stochastic processes literature, or can be derived by known methods. This program parallels that of Aldous [1] in which distributions associated with random trees are derived by coding trees as walks converging to Brownian excursion.

The exact way in which rBB approximates a random mapping is best said in pictures, but here is one aspect. Let $[G_1, D_1]$ be the excursion of Brownian bridge containing a uniform random time. Then we can decompose the Brownian bridge into three processes defined on $[0, G_1], [G_1, D_1]$ and $[D_1, 1]$, and these three processes are rescaled versions of Brownian bridge, Brownian excursion and Brownian bridge respectively. If we take a random mapping and see where vertex 1 is, we can split the graph into three parts:

(a) the tree-component containing 1;

(b) the rest of the graph-component containing 1;

(c) the rest of the graph.

In our coded walk these parts appear in order (b,a,c), and approximate the tripartite decomposition of rBB described above. This is joint work with Jim Pitman.

REFERENCES

1. D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Proc. Durham Symp. Stochastic Analysis 1990*, pp. 23-70, Cambridge University Press, 1991.

2. P. Flajolet and A. Odlyzko. Random mapping statistics. In J.-J. Quisquater, editor, *Proc. Eurocrypt '89*, pp. 329-354, Springer-Verlag, 1990. Lecture Notes in C.S. 434.

3. V.F. Kolchin. *Random Mappings*. Optimization Software, New York, 1986. (Translation of Russian original).

Search for the maximum of a random walk

A.M. ODLYZKO

AT&T Bell Laboratories

Let $X_1, X_2, \ldots$ be independent and identically distributed with $\mathsf{P}(X_j = 1) = \mathsf{P}(X_j = -1) = 1/2$, and let $S_k = X_1 + X_2 + \cdots + X_k$. Thus S_k is the position of a symmetric random walk on the line after k steps. Any algorithm that determines $\max\{S_0, \ldots, S_n\}$ with certainty must examine at least $c_1 n^{1/2}$ of the S_k on average for a certain constant $c_1 > 0$, if all random walks with n steps are considered likely. There is also an algorithm that on average examines only $c_2 n^{1/2}$ of the S_k to determine their maximum for another constant c_2. These results can be used to model some search problems on functions that are difficult to compute.

The height of a random partial order: concentration of measure

B. BOLLOBÁS

University of Cambridge, England

The problem of determining the length L_n of the longest increasing subsequence in a random permutation of $\{1, \ldots, n\}$ is equivalent to that of finding the height of a random 2-dimensional partial order (obtained by intersecting two random linear orders). The expectation of L_n is known to be about $2\sqrt{n}$. Frieze investigated the concentration of L_n about this mean, showing that, for $\alpha > \frac{1}{3}$, there is some constant $\beta > 0$ such that

$$\mathsf{P}(| L_n - EL_n | \geq n^{\alpha}) \leq \exp(-n^{\beta}). \tag{1}$$

In the talk we shall present some recent results obtained jointly with Graham Brightwell. It will be shown that (1) holds for all $\alpha > \frac{1}{4}$ as well, and analogous results are true for random k-dimensional orders, for each fixed $k \geq 2$.

Comparison techniques for card shuffling

P. DIACONIS

Harvard University

A new set of techniques has emerged to allow us to give useful answers to the following problem: given a set of permutations, create a random walk by repeatedly picking from this set (with replacement) and multiplying. How many steps does it take to get random? The method uses comparison techniques for Dirichlet forms. The basis for comparison is a well studied random walk where all the eigenvalues are known. The method seems to give sharp answers for almost any symmetric set of generators. In particular, it gives sharp results for the exclusion process. This is joint work with Laurent Salloff-Coste.

Long common subsequence problems

J.M. STEELE

University of Pennsylvania

If X_i and Y_i are independent random variables with values in the same alphabet, the variable L_n that we investigate is defined as the maximal m such that there are subsequences $i_1, i_2, \ldots, i_m$ and $j_1, j_2, \ldots, j_m$ of $\{1, 2, \ldots, n\}$ such that $X_{i_k} = Y_{j_k}$ for all $1 \leq k \leq m$. This talk briefly reviews recent progress on the tightness of concentration and other properties of this variable.

On menage problems

L. HOLST

Royal Institute of Technology, Stockholm

Consider n couples seated at circular tables with men and women taking alternating seats but otherwise completely random seating. Let W be the number of couples sitting next to each other. What can be said about W? That will be discussed especially when n is large.

Poisson limit laws for dependent random permutations

P.J. JOYCE

University of Idaho

The process of cycle counts $(C_1, C_2, \ldots, C_n, 0, 0, \ldots)$ for a random permutation Π of length n distributed according to the Ewens sampling formula converges to a Poisson process with independent coordinates. This result is extended to a vector of random permutations $\mathbf{\Pi} = (\Pi_1, \ldots, \Pi_d)$ in the following way. Let $\mathbf{Y} = (Y_1, \ldots, Y_d)$ be an integer valued random vector with $\sum_{i=1}^{d} Y_i = n$. Conditional on $\mathbf{Y}$, Π_i is a permutation of length Y_i distributed according to a Ewens sampling formula. For $i = 1, \ldots, d$ define $\mathbf{C}_i = (C_{i1}, \ldots, C_{in}, 0, \ldots)$, where C_{ij} is the number of cycles of size j in permutation Π_i. It can be shown that for a wide class of distributions for $\mathbf{Y}$, the $\mathbf{C}_i$ converge to independent Poisson processes. Total variation techniques are used to establish the result. The work is motivated by a problem in population genetics. This is joint work with Simon Tavaré.

On the structure of a non-uniformly distributed random graph

V.I. KHOKHLOV

Steklov Mathematical Institute, Moscow

We consider a random graph $G_{N,T}$ with N labelled vertices and T edges. These T edges are obtained by T independent trials. In each trial the edge between vertices i and j occurs with the probability $2p_ip_j$, and the loop at the vertex i occurs with the probability p_i^2; $i, j = 1, \ldots, n$, $p_1, \ldots, p_n \geq 0$, $p_1 + \ldots + p_n = 1$.

Let $N \to \infty$, $2T/N \to \lambda$, $p_i = a_i/N$, $a_i = a_i(N)$, $i = 1, 2, \ldots, N$, and there exists a limit

$$a^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} a_i^2$$

and positive constants ε and E such that $\varepsilon \leq a_i \leq E$, $i = 1, 2, \ldots, n$. Then, under the additional condition $\lambda a^2 < 1$, the graph $G_{N,T}$, with probability approaching one, does not contain components with more than one cycle and tree- components that have more than $c \log N$ vertices, where c is a constant. Moreover, under these conditions the distribution of the number of cycles in the graph converges to the Poisson distribution with parameter $\Lambda = -\frac{1}{2} \ln(1 - \lambda a^2)$.

REFERENCES

1. V. F. Kolchin and V. I. Khokhlov, On the number of cycles in a non-equiprobable random graph. *Diskretnaya Matematika* (1990) **2**, No. 3, 137–145 (in Russian).

2. V. I. Khokhlov and V. F. Kolchin, On the structure of a random graph with nonuniform distribution. In: *New Trends in Probab. and Statist.*, pp. 445–456, VSP/Mokslas, 1991.

The Poisson paradigm and random graphs

J. SPENCER

New York University

When a random variable X is the sum of many indicator random variables, each rare and mostly independent, the Poisson Paradigm is that X has close to the Poisson distribution. In particular - if $E[X] = \mu$ the $\Pr[X = 0]$ should be close to $e^{-\mu}$. This is a natural situation in Random Graphs. For example, in the original papers of Erdős and Rényi on Random Graphs it was shown that in $G(n,p)$ if $p = p(n)$ is such that the expected number of triangles is a constant μ then indeed the probability that there is no triangle approaches $e^{-\mu}$.

A few years ago Svante Janson, employing a variant of the Stein-Chen method, found a pair of inequalities now known as the Janson inequalities. With these results such as the above come out with a fairly elementary calculation on Random Graphs, involving no more, basically, than evaluation of the second moment. Applications include the following.

- Bounds on the probability that $G \sim G(n,p)$ contains no subgraph H – for a fixed H and various $p = p(n)$.
- Fine threshold behavior for every vertex to lie in a triangle, and similar extension statements.
- The existence of sets S of positive integers so that the number of representations $n = x + y + z$ with $x, y, z \in S$ is $\Theta(\ln n)$.

On the heights and widths of random rooted trees

L. TAKÁCS

Case Western Reserve University

We shall consider rooted trees with vertices $1, 2, \ldots, n$. The root is labeled 1. Each tree is represented by an ordered sequence of n nonegative integers $(i_1, i_2, \ldots, i_n)$ satisfying the conditions

$$i_1 + i_2 + \cdots + i_n = n - 1, \tag{1}$$

and

$$i_1 + i_2 + \cdots + i_r \geq r \text{ for } 1 \leq r < n. \tag{2}$$

Denote by S_n the set of all distinct trees defined above. In a tree, represented by $(i_1, i_2, \ldots, i_n)$, two vertices r and s $(1 \leq r < s \leq n)$ are joined by an edge if and only if

$$i_0 + i_1 + \cdots + i_{r-1} < s \leq i_0 + i_1 + \cdots + i_r \tag{3}$$

where $i_0 = 1$. The number of trees in S_n is

$$|S_n| = C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1}, \tag{4}$$

where $C_0 = C_1 = 1, C_2 = 2, C_3 = 5, C_4 = 14, \ldots$ are the Catalan numbers.

Let $\{p_j\}$ be a probability distribution on the set of nonnegative integers with expectation 1 and standard deviation σ $(0 < \sigma < \infty)$. Let us choose a tree at random in S_n, assuming that the probability of a tree represented by $(i_1, i_2, \ldots, i_n)$ is

$$p(i_1, i_2, \ldots, i_n) = A_n p_{i_1} p_{i_2} \ldots p_{i_n} \tag{5}$$

where A_n is determined by the requirement

$$\sum_{S_n} p(i_1, i_2, \ldots, i_n) = 1. \tag{6}$$

For a tree chosen at random in S_n define the random variable $\tau_n(m)$ as the number of vertices at distiance m from the root. Furthermore, define

$$\mu_n = \max\{m : \tau_n(m) > 0\} \tag{7}$$

as the height of the tree,

$$\delta_n = \max\{\tau_n(m) : m \geq 0\} \tag{8}$$

as the width of the tree and

$$\tau_n = \sum_{m\geq 0} m\tau_n(m) \tag{9}$$

as the total height of the tree.

Our aim is to find the asymtotic distributions of τ_n, μ_n, δ_n and $\tau_n(m)$ if $n \to \infty$ and $m = [2\alpha\sqrt{n}/\sigma]$ where $0 < \alpha < \infty$.

We shall prove the following theorems:

Theorem 1 *We have*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\sigma\tau_n}{\sqrt{4n^3}} \leq x\right\} = W(x) \tag{10}$$

for $x \geq 0$ where $W(x)$ is the distribution function of a positive random variable and is given by

$$W(x) = \frac{\sqrt{6}}{x}\sum_{k=1}^{\infty} e^{-\nu_k}\nu_k^{2/3}U(1/6, 4/3, \nu_k) \tag{11}$$

for $x > 0$. The function $U(a, b, x)$ is the confluent hypergeometric function,

$$\nu_k = 2a_k^3/(27x^2), \tag{12}$$

and $z = -a_k (k = 1, 2, \ldots)$ are the zeros of the Airy function $Ai(z)$ arranged so that $0 < a_1 < a_2 < \ldots < a_k < \ldots$.

Theorem 2 *If $0 < \alpha < \infty$, then*

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{2\tau_n([2\alpha\sqrt{n}/\sigma])}{\sigma\sqrt{n}} \leq x\right\} = G_\alpha(x) \tag{13}$$

for $x > 0$ where $G_\alpha(x)$ is the distribution function of a nonnegative random variable and is given by

$$G_\alpha(x) = 1 - 2\sum_{j=1}^{\infty}\sum_{k=0}^{j-1}\binom{j-1}{k}e^{-(x+2\alpha j)^2/2}(-x)^k H_{k+2}(x + 2\alpha j)/k! \tag{14}$$

for $x \geq 0$ where $H_0(x), H_1(x), \ldots$ are the Hermite polynomials defined by

$$H_n(x) = r\,! \sum_{j=0}^{[n/2]} \frac{(-1)^j x^{n-2j}}{2^j j!(n-2j)!}. \tag{15}$$

Furthermore, we have

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\sigma\mu_n}{2\sqrt{n}} \leq x\right\} = \lim_{n\to\infty} \mathbf{P}\left\{\frac{\delta_n}{\sigma\sqrt{n}} \leq x\right\} = G_x(0) \tag{16}$$

for $x > 0$. The first part of this equation is due to V.F. Kolchin (1978), and the second part was conjectured by D. Aldous (1990).

Let $\{\eta^+(t), 0 \leq t \leq 1\}$ be the Brownian excursion process and $\tau^+(\alpha)$ its local time at level $\alpha \geq 0$. Let

$$\tau^+ = \int_0^1 \eta^+(t)dt. \tag{17}$$

The aforementioned results for random trees imply that

$$\mathbf{P}\{\tau^+(\alpha) \leq x\} = G_\alpha(x) \tag{18}$$

for $\alpha > 0$ and $x > 0$ and

$$\mathbf{P}\{\tau^+ \leq x\} = W(x) \tag{19}$$

for $x \geq 0$. It is already known that

$$\mathbf{P}\left\{\sup_{0\leq t\leq 1} \eta^+(t) \leq x\right\} = \mathbf{P}\left\{\max_{\alpha\geq 0} \tau^+(\alpha) \leq 2x\right\} = G_x(0) \tag{20}$$

for $x \geq 0$. [D.P. Kennedy (1976) and T. Jeulin (1985).]

Random permutations, limit shapes and asymptotic problems of partition theory

A. VERSHIK

St. Petersburg University

Asymptotic properties of the limit measures which appear in additive problems in partition theory, and number theory can be considered in some geometric manner. There are the different type of asymptotic behaviour: 'ergodic' in which we can put the problems like LLN and CLT and 'nonergodic' in which one can calculate the limit distribution and the boundary. All kind of these examples appear in the context of random permutations and statistics on the partitions on integers or reals. It happens that completely different problems can give us the same limit measures.

In search of maximum antichains of partitions

L.H. HARPER

University of California, Riverside

An antichain in a poset P is a subset of P having no comparable members. In 1928 Sperner showed that the largest antichain in the set of all subsets of an n-set ordered by containment, is its largest rank.

Around 1967 Rota asked if the partitions of an n-set, ordered by refinement, has the Sperner property, i.e. is the largest antichain the largest rank? Erdős has asked the same question about the partitions of n.

In 1978 Canfield showed that the answer to Rota's question is 'no'; that for n sufficiently large ($n > 6 \times 10^{24}$) antichains larger than any rank exist. Subsequent papers by Shearer and Kleitman lowered the upper bound on the Canfield number to 6×10^6, but did not give any lower bounds on it nor ascertain whether there were antichains significantly larger than the largest rank.

In 1985 the present author

i) Showed how to approximate the poset of partitions of an n-set by a Gaussian process ordered by a cone.

ii) Solved the finite dimensional analog of the Sperner problem in i);

iii) Carried out calculations, based on the assumption that the solution of the Sperner problem is preserved by the limiting process of i), which show that the Canfield number is about 6×10^6 and that the ratio of the largest antichain to the largest rank converges to 1.69.

Recently J. Chavez and I have been investigating whether the same technique can answer Erdős's question. Our latest results will be presented.

REFERENCES

1. E.R. Canfield, On a problem of Rota, *Advances in Math.* (1978) **29**, 1–10.

2. L.H. Harper, On a continuous analog of Sperner's problem, *Pacific J. Math.* (1985) **118**, 411–425.

3. G.C. Rota, Research problem: A generalization of Sperner's theorem, *J. Combinatorial Theory* (1967) **2**, 104.

4. J.B. Shearer, A simple counterexample to a conjecture of Rota, *Discrete Math.* (1979) **28**, 327–330.

5. E. Sperner, Ein Satz uber Untermengen einer endlichen Menge, *Math Z.* (1928) **27**, 544–548

On sets of integers with prescribed gaps

W. STADJE

University of Osnabrück, Germany

For a fixed set I of positive integers we consider the set of paths $(p_0, ..., p_k)$ of arbitrary length satisfying $p_l - p_{l-1} \in I$ for $l = 2, ..., k$ and $p_0 = 1, p_k = n$. Equipping it with the uniform distribution, the random path length T_n is studied. Asymptotic expansions of the moments of T_n are derived and its asymptotic normality is proved. The step lengths $p_l - p_{l-1}$ are seen to follow asymptotically a restricted geometrical distribution. Analogous results are given for the free boundary case in which the values of p_0 and p_k are not specified. In the special case $I = \{m+1, m+2, ...\}$ (for some fixed $m \in \mathbb{N}$) we derive the exact distribution of a random 'm-gap' subset of $\{1, ..., n\}$ and exhibit some connections to the theory of representations of natural numbers. A simple mechanism for generating a random m-gap subset is also presented. This is joint work with Y. Baryshnikov.

On random partitions of the integer n

E. SCHMUTZ

Drexel University

Put a uniform probability distribution on the set of partitions of the integer n into parts that are elements of a certain set A. If $A_1, A_2, \ldots, A_d$ are disjoint sets whose union is A, let $P_i(\lambda)$ denote the number of part sizes that the partition λ has in A_i. Under suitable conditions, the random vector $P = (P_1, P_2, \ldots, P_d)$ is asymptotically normally distributed.

Random partitions of large integers

B. FRISTEDT

University of Minnesota

Random partitions of integers will be discussed for the case where all partitions of an integer n are equally likely. The focus is on limit theorems as $n \to \infty$. In particular, as $n \to \infty$, the decreasing sequence of large parts, beginning with the largest part, then the next largest part, etc. approaches, when appropriately normalized, a certain Markov chain which can be explicitly identified. The major tool is a simple construction of random partitions that treats the number being partitioned as a random variable in such a way that the numbers of parts of various sizes are independent random variables.

Sampling properties of random mappings

W.J. EWENS

University of Pennsylvania

Three aspects of 'sampling' in a random mapping from $(1, 2, \ldots, m)$ to $(1, 2, \ldots, m)$ will be considered. In the first of these, we consider the components of $(1, 2, \ldots, n)$ induced by the components of the mapping. We find that often parameters do not converge from the sample to the original mapping: this is due to our inability to use weak convergence theory. The second concerns the relation between samples from a random mapping and a random permutation: as m increases without limit, the two distributions become identical, apart from the value of a certain parameter. The third concerns size-biased sampling: the components of the mapping are listed by a size-biased procedure. The limiting distribution of the normalized component sizes converges to a distribution with many remarkable properties. The distributions described above arise often in population genetics theory, and their interpretation in that field will be discussed.

The height of elements in random mappings

D.R. GAVELEK

XonTech Inc.

Interest in the properties of random mappings - arbitrarily selected from the set of N^N functions that map a set of N distinct elements into itself - was stimulated almost forty years ago by Metropolis and Ulam. Over the intervening years random mapping models have been used in applications ranging from random number generation and cryptography to the simulation of epidemic processes and tests of the intrinsic randomness of quantum mechanics. Discussed in this paper are approximately twenty characteristics related to the height distributions of elements in the functional graphs or de Bruijn diagrams representing a random map. Some of these distributions are well known. Other results, such as the expected number of ancestors of an element of height H, and the average height of an orphan point, appear to be new. As an additional unifying factor it is shown that all of these parameters are naturally expressed in terms of the incomplete gamma function.

The evolution of a random mapping

J. JAWORSKI

Adam Mickiewicz University, Poland

A random mapping $(T_n; q)$ of a finite set $V = \{1, \ldots, n\}$ into itself assigns independently to each $i \in V$ its unique image $j = T(i) \in V$ with probability q for $i = j$ and with probability $P = (1-q)/(n-1)$ for $i \neq j$. We study the evolution of a random digraph $G_{T_n}(q)$, representing $(T_n; q)$, as its arc-occurrence probability $P = P(n)$ increases from 0 to $1/(n-1)$. The structure of functional digraphs enables asymptotic studies of exact discrete distributions of many characteristics related to G_{T_n}. For example, we consider the number of predecessors of m given vertices, the quasi-binomially distributed random variable associated with a particular epidemic process. Finally, let $(T_n; M)$ be a random element of a family of all loopless digraphs on n vertices with exactly M vertices of outdegree 1 and $n - M$ vertices of outdegree 0. Clearly, there is an equivalence between $(T_n; q)$ and $(T_n; M)$. Moreover, $(T_n; M)$ can be treated as the Mth stage of the 'regular' random digraph process $\{G_{T_n}(M)\}_{M=0}^{n(n-1)}$. We study the appearance of the first cycle in such a process and the structure of the digraph near the critical point $M = n$.

Order statistics for random combinatorial structures

J.C. HANSEN

Northeastern University

We consider labeled and unlabeled 'decomposable' combinatorial structures which are characterized by the following generating function equations. In the labeled case, $\hat{P}(z) = \exp \hat{C}(z)$ where $\hat{P}(z)$ is the exponential generating function for the number of structures of size n and $\hat{C}(z)$ is the exponential generating function for the number of connected structures of size n. In the unlabeled case, $P(z) = \exp(C(z) + C(z^2)/2 + \ldots)$ where $P(z)$ is the ordinary generating function for the number of structures of size n and $C(z)$ is the ordinary generating function for the number of 'connected' structures of size n. In both cases, we are interested in the measure induced $\nabla = \{\{x_i\} : x_1 \geq x_2 \geq \ldots \geq 0, \sum x_i \leq 1\}$ by the (decreasing) sequence of order statistics for the component sizes of a random structure of size n (normalized by n). We show that if the generating functions $\hat{C}(z)$, in the labeled case, and $C(z)$, in the unlabeled case, are logarithmic functions then the induced measures on ∇ converge in distribution to a Poisson-Dirichlet distribution on ∇. In the labeled case, this result unifies results known for particular examples such as random permutations and random mappings. In the unlabeled case, this gives new distributional results for examples such as factorization of polynomials over $GF(q)$.

The Rota umbral calculus and the Heisenberg-Weyl algebra

O.V.VISKOV

Steklov Mathematical Institute, Moscow

The talk is based on some of the author's papers (see [1] and references therein, [2], [3]).

The Heisenberg-Weyl algebra is the algebra freely generated by three variables A,B and C subject to the identities

$$[A, B] = AB - BA, \qquad [A, C] = [B, C] = 0.$$

The main purpose of this talk is to emphasize the role played by a suitable representation of this algebra in contemporary umbral or operator calculus originated in an inspiring paper of G.-C. Rota [4].

Let $p = \{p_n(x), n = 0, 1, 2, \ldots\}$ be an arbitrary basis in the commutative algebra $\mathcal{P}$ of all polynomials of a single variable x with coefficients in a field of characteristic zero and let $\mathcal{L}$ be the set of linear maps from $\mathcal{P}$ into $\mathcal{P}$. Since every operator in $\mathcal{L}$ is uniquely determined by its actions on an arbitrary basis p of $\mathcal{P}$, the relations

$$A[p_0(x)] = 0, \qquad A[p_n(x)] = np_{n-1}(x), \qquad n = 1, 2, \ldots;$$

$$B[p_n(x)] = p_{n+1}(x), \qquad n = 0, 1, 2, \ldots,$$

give us the desirable representation of the Heisenberg Weyl algebra if we take the identity map as C. The simplest particular case of this situation is $p_n(x) = x^n$, $n = 0, 1, 2, \ldots$, and then $A = \mathrm{d/dx}$ and the operator B is the multiplication by x.

Representation (2) and the systematic use of relations (1) allow us to obtain easily many important formulae in the finite operator calculus [5], [6] including the recurrence and transfer formulae, umbral operators and effective tool for the composition and inversion of power series. Moreover, the proofs become essentially simpler.

The above mentioned approach admits a simple generalization to the multivariate case. It is also useful for analysis of many other situations, for instance, in the Stanley theory of differential posets [7].

REFERENCES

1 O. V. Viskov. A noncommutative approach to classical problems of analysis. *Trudy Mat. Inst. Steklov* (1986) **177**, 21–32. (Russian) English transl. in Proc. Steklov Inst. Math. 1988, issue 4, 21–32.

2. O.V. Viskov. A class of linear operators. In *Generalized functions and their applications*, Proc. Inter. Conference, Moscow, 1981,pp. 110–120 (Russian).

3. O.V. Viskov. On the ordered form of noncommutative binomial. *Uspechi Mat. Nauk* (1991) **46**, 209–210. (Russian).

4. G.-C. Rota. The number of partitions of a set. *Amer. Math. Monthly* (1964) **71**, 498 - 504.

5. G.-C. Rota (1975) *Finite Operator Calculus* Academic Press, New York.

6. S.M. Roman. (1984) *The Umbral Calculus.* Academic Press, Orlando, FL.

7. R.P. Stanley. Differential posets. *J. Amer. Math. Soc.* (1988) **1**, 919 - 961.

Random permutations and stable matchings

B. G. PITTEL

Ohio State University

A matching on a set of an even number of members is stable - with regard to a given system of members' preferences for a partner - if no two unmatched members prefer each other to their partners under the matching. We study the set of stable matchings for a random instance of the ranking system, under an assumption that each member rank orders potential partners uniformly at random, independently of other members. For the bipartite version ("stable marriages") with n men and n women, we prove that almost surely the total number of stable matchings (marriages) is at least $(n/\log n)^{1/2}$. We show an almost sure (a.s.) existence of an "egalitarian" marriage, for which the total rank of all spouses is about $2n^{3/2}$, as opposed to $n^2/\log n$ for the extreme - female optimal and male optimal - marriages. A.s. this particular matching is also (asymptotically) a "minimum regret" stable marriage, with the largest rank of a spouse in it being close to $n^{1/2}\log n$. Quite unexpectedly, the stable matchings obey - statistically - a law of hyperbola. It states that almost surely the product of the sum of husbands' ranks and the sum of of wives' ranks in a stable marriage is asymptotic to n^3, *uniformly* over all stable marriages.

We also study a nonbipartite version of the stable matching problem, which is colloquially known as a "stable roommates" problem. Here, in a set of even cardinality n, each member ranks all others in order of preference. It is well known that unlike the bipartite version (marriages), a stable matching may not exist. We prove that, for the random instance of the ranking system, the mean and variance of the number of stable matchings are asymptotic to $e^{1/2}$ and $(\pi n/4e)^{1/2}$, respectively. (For the marriages, the mean is about $n\log n/e$.) Thus, $P(n)$ the probability that a solution exists is at least $const/n^{1/2}$. What is lim P(n)? Rob Irving has performed extensive computer runs using his two-stage proposal algorithm. (The algorithm delivers a stable matching whenever there is one.) The empirical data has lead him to an intriguing conjecture that the limit is positive. We present some preliminary results concerning the likely behavior of Irving's algorithm, and a hyperbola law which holds for stable tables, Irving's extension - to the nonbipartite

case of the notion of stable marriages. We also discuss likely structure of closely related stable (cyclic) partitions, which have been discovered recently by Tan.

Finally, we look at the probabilistic aspects of a conceptually (and mathematically) related trade model introduced and investigated by Shapley and Scarf.

Independent process approximations for combinatorial structures

R.A. ARRATIA

University of Southern California

Many random combinatorial objects have a structure whose joint distribution is exactly equal to that of a process of mutually independent random variables, conditioned on the value of a weighted sum of these independent random variables. It is interesting to compare the combinatorial structure directly to the independent process, without conditioning. The quality of approximation can be conveniently quantified in terms of total variation distance. Examples include random permutations, random mapping functions and patterns, random partitions of a set, random partitions of a positive integer, random partitions of an integer into parts of distinct sizes, and random polynomials over a finite field, in every case with the uniform distribution over all possibilities of size n.

In more detail, consider the component structure $\mathbf{C}(n) = (C_1(n), C_2(n), \ldots, C_n(n))$, where C_i represents the number of parts of size i. The random variables $C_1, C_2, \ldots, C_n$ are mutually dependent, since the weighted sum $C_1 + 2C_2 + \cdots + nC_n$ has the constant value n. For a given family of combinatorial objects, indexed by $n = 1, 2, \ldots$, and for each value of a real parameter $x > 0$, there are mutually independent random variables $Z_1, Z_2, \ldots$, with the following property. Let T_n be the weighted sum $T_n = Z_1 + 2Z_2 + \cdots + nZ_n$. For each n, the joint distribution of the combinatorial process $\mathbf{C}(n)$ is equal to the joint distribution of $(Z_1, \ldots, Z_n)$, conditioned on the event $\{T_n = n\}$. The simple, independent process $(Z_1, \ldots, Z_n)$, without conditioning on the value of T_n, may directly provide useful approximations to the distribution of $\mathbf{C}(n)$.

To describe the independent random variables Z_i, let m_i be the number of possible structures available for each part of size i. For the class of combinatorial assemblies, which includes permutations, with $m_i = (i-1)!$, mapping functions, with $m_i = (i-1)!\,(1 + i + i^2/2 + \cdots + i^{i-1}/(i-1)!)$, and partitions of a set, with $m_i = 1$, we have that Z_i is Poisson with parameter $\lambda_i = m_i x^i / i!$, $x > 0$. For the class of multisets, which includes partitions of an integer, with $m_i = 1$, polynomials of degree n, with $m_i =$ the number of monic irreducible polynomials of degree i, and random mapping patterns, we have that Z_i is negative binomial, corresponding to the sum of m_i independent geometric random variables Y with $\mathsf{P}(Y = k) = (1 - x^i)(x^i)^k$ for $k \geq 0, 0 < x < 1$. For the class of selections, which includes partitions of an integer into distinct parts, and square free polynomials, we have that Z_i is Binomial with parameters $m_i, x^i/(1 + x^i), x > 0$.

An appropriate choice of the parameter x corresponds roughly to maximizing $\mathsf{P}(T_n = n)$. In some cases, a constant gives an appropriate choice of x; examples include $x = 1$ for random permutations, $x = e^{-1}$ for random mapping functions, $x = \rho \equiv 0.3383\ldots$ for random mapping patterns, and $x = q^{-1}$ for random polynomials over a field with q elements. In such cases, $C(n)$, viewed as an element of $\mathbf{R}^\infty$, converges in distribution to $(Z_1, Z_2, \ldots)$. In other examples an appropriate choice of x must vary with n; examples include $x = \log n - \log\log n$ for random partitions of the set $\{1, 2, \cdots, n\}$, $x = \exp(-\pi/\sqrt{6n})$ for random partitions of the integer n and $x = \exp(-\pi/\sqrt{12n})$ for partitions with no repeated parts.

For most examples, with an appropriate choice of x, for large n and individually for each $i = 1$ to n, Z_i is a good approximation for $C_i(n)$. More generally, for $B \subset \{1, \ldots, n\}$, the joint distribution of the independent process $(Z_i)_{i \in B}$ is a good approximation for the joint distribution of the process $(C_i)_{i \in B}$, provided that B is small in the sense that the contributions to the mean and variance of T_n from terms indexed by B are small compared to the mean and variance of T_n. This approximation can be quantified conveniently by the total variation metric, and allows effective approximation of the distribution of some functionals of the entire process $\mathbf{C}(n)$ by the distribution of the same functional applied to the independent process $(Z_1, \ldots, Z_n)$. Clearly, not all functionals are approximated well in distribution, the extreme example being the indicator functional $h(a_1, \ldots, a_n) = 1\{a_1 + 2a_2 + \cdots + na_n = n\}$, since in all non-trivial examples $\mathsf{E}h(Z_1, \ldots, Z_n) = \mathsf{P}(T_n = n) \to 0$.

We consider issues common to all the above examples, including equalities and upper bounds for total variation distances, heuristics for good approximations, the relation to standard generating functions, refinement to the process which counts the number of parts of each possible type, the effect of conditioning on further restrictions, large deviation theory and nonuniform measures on combinatorial objects, and the possibility of getting useful upper bounds for the probability of unlikely events by simply giving a lower bound on $\mathsf{P}(T_n = n)$. Detailed examples, which show the utility and tractability of these approximations of combinatorial processes by independent processes, appear in separate papers. This is joint work with Simon Tavaré.

Linear and non-linear codes for a special channel

L.A. SHEPP

AT&T Bell Laboratories

Janos Korner et al want to construct a subset, S, of the group $G_n = \{0, 1, -1\}^n$ with componentwise addition modulo 3, with as many elements as possible and with the property that any two code words (elements of S) x, y are far apart in the sense that some component of the difference, is 1 ie $x(i) - y(i) = 1$ for some i. (It follows that some other component of $x - y$ is -1, by interchanging the roles of x and y.)

Let $A(n)$ be the cardinality of any set S attaining maximum cardinality with the property. It is easy to see that $A(m+n) \geq A(m)A(n)$ and so $A(n) \approx a^n$ for $n \to \infty$ for some a, where $2 \leq a \leq 3$. The lower bound comes from the example, due to the proposers, of the set

$$S = \{x : x \text{ has } n/2\ 1's \text{ and the rest } 0's\}.$$

If S is also required to be a subgroup of G_n, then the maximum cardinality will be a number, $B(n)$, and again $B(n) \approx b^n$ as $n \to \infty$ for some b, where $\sqrt{3} \leq b \leq a$. The lower bound comes from the example

$$S = \{c_1(1, -1, 0, 0, \ldots) + c_2(0, 0, 1, -1, 0, 0, \ldots) + c_3(0, 0, 0, 0, 1, -1, 0, 0, \ldots) + \ldots\}$$

of $n/2$ generators of the subgroup S with $n/2$ independent coefficients $c_1, c_2, \ldots$ in $\{0, 1, -1\}$.

We prove that these lower bounds are asymptotically best possible (for $A(n)$, only in the weak sense of $\approx$), so that $b = \sqrt{3}$, and $a = 2$. Thus there are considerably fewer codewords possible if S is required to be a linear, or group code, which seems to contradict, to some extent, the naive belief that the more efficient (in the sense of decodability) group codes give up little in the sense of capacity. This is joint work with Rob Calderbank, Peter Frankl, Ron Graham, and Wen-Ch'ing Winnie Li.

A lower bound on the probability of conflict under non-uniform access in database systems

P. MATTHEWS

University of Maryland Baltimore County

We consider an N item database and t transactions, each of which will independently request a subset of the items. For $i = 1, \ldots, t$ transaction i will request n_i items according to some probability distribution on the $\binom{N}{n_i}$ sets of n_i items. We say there are no conflicts if the t chosen sets are all disjoint. In probabilistic language t complexes of balls are being allocated independently to N urns, and we are considering the probability that no urn receives two or more balls. If the transactions choose simple random samples, then the probability of no conflicts is $\binom{N}{n_1, \ldots, n_t} \left(\prod_{i=1}^{t} \binom{N}{n_i} \right)^{-1}$. We give a class of sampling schemes of practical interest and show that, within this class, the probability of no conflicts is no larger than that for simple random sampling. This supports a long-standing conjecture in the database community that uniform access minimizes the probability of conflicts. This is joint work with Keith Humenik, A. B. Stephens, and Yelena Yesha.

Covering with Latin Transversals

P. TETALI

DIMACS center, Rutgers University,

Given an $n \times n$ matrix $A = [a_{ij}]$, a *transversal* of A is a set of elements, one from each row and one from each column. A transversal is a *Latin transversal* if no two elements are the same. There have been more conjectures than theorems on latin transversals in the literature. Recently, Erdős and Spencer showed that there always exists a latin transversal in any $n \times n$ matrix in which no element appears more than s times, for $s \leq (n-1)/16$. Here we show that, in fact, all the elements of the matrix can be partitioned into latin transversals, provided n is a power of 2 and no element appears more than ϵn times for some fixed $\epsilon > 0$.

Theorem *Let n be 2^m. Any $n \times n$ matrix in which no element appears more than s times contains n disjoint latin transversals provided $s \leq \epsilon n$ (for ϵ, an absolute constant $\ll 1$).* The assumption that n is a power of 2 can be weakened, but at the moment we are unable to prove the theorem for all values of n. On the other hand, our proof can be easily modified to prove the existence of many pairwise disjoint transversals in any n by n matrix in which no entry appears more than ϵn times, without any restriction on n. Therefore our method implies a strengthening of the result of Erdős and Spencer for any n, (apart from the actual value of the constant ϵ). The proof of Theorem 1 involves *random partitioning* and *Lovász local lemma.* This is joint work with N. Alon and J. Spencer.

On $([n], P)$-Partitions

Z.-X. HU

University of Illinois

In general, for a set X, probably two of the most interesting structures X has are order relations and equivalent classes. We introduce the concept of $([n], P)$-partitions which is the combination of the above two structures (It is inspired by Richard P. Stanley's interesting (P, ω)-partitions), where P is a poset and $[n] = \{1, 2, ..., n\}$. Many mathematical models can be considered as special cases of $([n], P)$-partitions. One interesting application of $([n], P)$-partitions is the special way of realizing a finite poset P by the complete graph K_n on $[n]$. A new parameter $n(P)$ (called the norm of P) is naturally derived. n(P) gives an indication of the complexity of P by reflecting both $|P|$ and order relations in P. We will introduce some new results about $([n], P)$-partitions and $n(P)$.

Some results on Poisson and compound Poisson approximation

A.P. GODBOLE

Michigan Technological University

The probability density function for the number of overlapping (or non-overlapping) occurrences of a word pattern can be expressed in terms of multinomial coefficients summed over an index set determined by an integer partition. We will present Poisson and compound Poisson approximations for such random variables, given i.i.d. or stationary Markov letter generation. The techniques employed include univariate and process versions of the Stein-Chen method (due to Arratia, Goldstein, Gordon, Barbour, Holst and Janson) and eigenvalue bounds on convergence to stationarity for a non-reversible Markov chain (due to Diaconis, Stroock and Fill). Different compound Poisson approximations will be provided for the number of matches (within a finite memory window of size k) while sampling with replacement, and for a non-i.i.d. urn problem related to the determination of the number X of winners of a lottery jackpot.

Labellings, size-biased permutations and the GEM distribution

P.J. DONNELLY

Queen Mary and Westfield College, London

In proving limit theorems for the 'sizes' of 'components' of combinatorial objects there are usually several ways of labelling the components. One labelling is by decreasing size order, another is a particular random labelling called a size-biased permutation. Continuity results usually guarantee that convergence with one labelling is equivalent to convergence with the other. In many cases (random permutations, random mappings, population genetics, prime divisors) normalised sizes converge to the Poisson-Dirichlet distribution with the ordered labelling and to the GEM distribution with the size biased labelling. Apart from its inherent interest and natural interpretation in some settings, use of the size-biased permutation often greatly facilitates the proof of convergence results. Furthermore, in stark contrast to the Poisson-Dirichlet, the GEM distribution which arises in the limit is extremely tractable. We also discuss consistency properties of the two distributions, and of samples from them.

Branching processes with final types of particles and random trees

V.A. VATUTIN

Steklov Mathematical Institute, Moscow

A natural analogy between some properties of branching processes and random trees firstly noted (in non-explicit form) by Otter [1] was rediscovered and widely used many times (see, for example, [2–5]). My intention to look at this analogy once more was stimulated by the last two of these papers. Kolchin's work [4], in which the analogy mentioned above obtained a rigorous mathematical description, considers the problem of finding the limit distribution, as $N \to \infty$, of the height of a tree chosen at random from the set of all rooted labelled trees having N vertices and edges of unit length. Article [5] investigates the same problem for a tree constructed in a slightly more complicated way: firstly one chooses at random a tree from the set of all binary trees having N vertices of degree 1 and, secondly, independent exponentially distributed random lengths are assigned to the edges of the tree. The methods used in these papers to find desired limit distributions are quite different. Although each of these methods rests upon a correspondence between the trees in question and some Markov branching processes, the first of them uses essentially some local limit theorems while the second one employs the moments method.

It appears that some modification of Kolchin's method can be used to prove not only the results from [5] but also more general statements. The key idea is to use a Bellman–Harris branching process with two types of particles one of which is final.

The process in question is described as follows. It is initiated at time $t = 0$ by the birth of one particle of non-final type which lives for a random time l with the distribution function $G(t) = \mathsf{P}\{l \le t\}$ and at the end of its life produces a random number ξ of non-final particles with the generating function $f(s) = \mathsf{E}s^{\xi}$ and, besides this, one particle of final type if $\xi \in \Delta$ where Δ is some fixed subset of non-negative integers. The final particle (if any) does not change after its birth moment while the newborn particles of non-final type evolve independently and stochastically the same as the initial one: each of them lives for a random time and at the end of its life produces particles of final and non-final types, and so on.

Let $z(t)$ be the number of non-final particles in the process at time t, $\tau = \inf\{t > 0\colon\ z(t) = 0\}$ and let ν_Δ be the total number of particles of final type which were born in the process up to the moment τ.

THEOREM. *Let* $p = \sum_{k\in\Delta} \mathsf{P}\{\xi = k\} > 0$, *and the G.C.D. of those* $k \in \Delta$ *for which* $\mathsf{P}\{\xi = k\} > 0$ *be 1. If* $\mathsf{E}\xi = 1$, $0 < f''(1) = B < \infty$ *and*

$$1 - G(t) = o(t^{-4}), \qquad t \to \infty,$$

then

$$\lim_{N\to\infty} \mathsf{P}\left\{ \frac{\tau}{\mathsf{E}l}\sqrt{\frac{B}{N}} \le x \,\middle|\, \nu_\Delta = N \right\} = 1 + 2\sum_{k=1}^{\infty}(1 - pk^2x^2)\, e^{-pk^2x^2/2}.$$

From the theorem, letting $f(s) = e^{s-1}$, $G(t) = 0$ if $t < 1$, $G(t) = 1$ if $t \ge 1$, $\Delta = \{0, 1, 2, \ldots\}$ and $p = 1$ we obtain, using the correspondence established in [4], Kolchin's theorem concerning the height of a random labelled rooted tree [4]; letting $f(s) = \frac{1}{2} + \frac{1}{2}s^2$, $G(t) = 1 - e^{\lambda t}$ if $t > 0$, $G(t) = 0$ otherwise, and $\Delta = \{0\}$, $p = 1/2$, we can arrive after some additional arguments to the result due to Gupta et al. [5], and finally, letting $f(s) = (2 - s)^{-1}$, $G(t) = 0$ if $t < 1$, $G(t) = 1$ if $t \ge 1$, $\Delta = \{0, 1, 2, \ldots\}$, $p = 1$ we can obtain the limit distribution of the height of a planted plane tree chosen at random from the set of all planted plane trees having N vertices firstly given in [6].

Some other problems closely connected with that described above will be considered also.

REFERENCES

1. R. Otter, The number of trees. *Ann. Math.* (1948) **49**, 583–589.

2. V. E. Stepanov, Limit distributions of certain characteristics of random mappings. *Theory Probab. Appl.* (1969) **14**, 612–626.

3. D. P. Kennedy, The Galton–Watson process conditioned on the total progeny. *J. Appl. Prob.* (1975) **12**, 800–806.

4. V. F. Kolchin, Moment of degeneration of a branching process and height of a random tree. *Math. Notes* (1978) **6**, 954–961.

5. V. K. Gupta, O. J. Mesa and E. Waymire, Tree-dependent extreme values: the exponential case. *J. Appl. Prob.* (1990) **27**, 124–133.

6. Yu. L. Pavlov, Some properties of planted plane trees. In: *Discrete Mathematics and Its Application to the Modelling of Complex Systems: Abstracts of the All-Union School.* Irkutsk, 1991, p. 14.

Refined approximations for the Ewens sampling formula

A.D. BARBOUR

University of Zürich, Switzerland

The Ewens sampling formula is a family of probability distributions over the space of cycle types of permutations of n objects, indexed by a real parameter θ. In the case $\theta = 1$, where the distribution reduces to that induced by the uniform distribution on all permutations, the joint distribution of the number of cycles of lengths less than $b = o(n)$ is extremely well approximated by a product of Poisson distributions, having mean $1/j$ for cycle length j: the error is super-exponentially small with nb^{-1}. For $\theta \neq 1$, the analogous approximation, with means adjusted to θ/j, is good, but with error only linear in $n^{-1}b$. In this paper, it is shown that, by choosing the means of the Poisson distributions more carefully, an error quadratic in $n^{-1}b$ can be achieved, and that essentially nothing better is possible.

Cycles and descents of random permutations

J. W. PITMAN

University of California, Berkeley

Formulae for the joint distribution of the cycle structure and number of descents of a random permutation are derived from simpler formulae for the distribution of the cycle structure of certain random riffle shuffles with at most $a-1$ descents. The results for the cycle structure of riffle shuffles assume a product form parallel to classical results for uniform random permutations, and involve the number of aperiodic circular words of a letters (or necklaces of a colors) of length n. This is joint work with Persi Diaconis and Michael McGrath.